doi:10. 13866/j. azr. 2019. 04. 17

基于 Sentinel－2 数据的干旱区典型绿洲植被叶绿素含量估算①

顾　峰[1,2,3]，　丁建丽[1,2,3]，　葛翔宇[1,2,3]，　高石宝[4]，　王敬哲[1,2,3]

（1. 新疆大学资源与环境科学学院，新疆 乌鲁木齐　830046；
2. 新疆大学智慧城市与环境建模自治区普通高校重点实验室，新疆 乌鲁木齐　830046；
3. 新疆大学绿洲生态教育部重点实验室，新疆 乌鲁木齐　830046；　4. 浙江大学地球科学学院，浙江 杭州　310027）

摘　要：以渭干河—库车河绿洲（渭—库绿洲）为研究区，采用在机器学习方面具有明显优势的随机森林回归算法，对绿洲内的4种典型植被（棉花、芦苇、杨树、大枣）叶片的叶绿素相对含量（soil and plant analyzer development，SPAD）进行估算和验证。首先基于“红边”处光谱信息丰富的哨兵2号（Sentinel－2）影像和由其衍生的一阶微分、二阶微分影像各提取23种对叶绿素敏感的宽波段光谱指数，加入3种影响植物生长的土壤参量（土壤含水量，土壤有机质，土壤电导率）作为影响叶片SPAD的特征变量，再根据以上特征变量对每种植被叶片各建立3种方案的SPAD估算模型，从而实现对绿洲内植被叶绿素的监测。结果表明：① 影像经一阶微分再提取的植被指数相比原位光谱植被指数，在SPAD估测模型中起到了更重要的作用，在随机森林算法的重要性排序中位居前列；② 4种植被叶片的SPAD估测模型都取得了不错的效果，芦苇叶片尤为显著，确定系数（R^2）达到了0.926；③ 分析对比3种方案下模型预测能力，方案3（包含土壤参量）的预测能力卓越〔2.143 < 相对百分比偏差（*RPD*）< 2.692〕，其预测能力排序为：方案3 > 方案1 > 方案2，土壤属性和模型预测结果有较强的非线性相关。Sentinel－2数据具有理想的估算绿洲植被叶绿素含量的潜力，提供了一种高效、低成本、潜在高精度的方案来估算叶绿素含量，可为干旱区绿洲农业、生态系统实现更有效的保护和管理提供参考。

关键词：绿洲；Sentinel－2数据；SPAD；叶绿素；植被指数；随机森林；新疆

叶绿素是光合作用过程中吸收光能的植被色素，更是植被（尤其在衰老期）光合作用能力的指征，已成为监测绿洲植被生长健康状况的重要的生理参数[1-2]。大量研究表明植被叶片的叶绿素相对含量（SPAD）可直接用以衡量其实际叶绿素含量的高低[3-7]。植被是绿洲生态系统的重要组成分，获取SPAD这一植被典型的生化参数并进行系统性研究，可以及时掌握绿洲生态系统的发展动向[8]。干旱区绿洲植被叶片叶绿素含量及时准确地估算及诊断，对制定科学的水肥管理措施及进行干旱状况评估具有重要的现实意义。受叶片中叶绿素吸收作用的影响，农作物叶片在“红边”位置（690～780 nm）存在着明显的光谱响应，这使得利用遥感技术对SPAD进行遥感估算成为了可能。常见的“红边”参数包括：红边位置、红边幅值、红边峰值等[9]。目前，MODIS、Landsat、HJ－1、ASTER、QuickBird等多光谱遥感数据已成功应用于植被理化参数的监测[10-13]，但它们通常缺少“红边”信息，这对植被SPAD的遥感估算是极为不利的。欧洲太空局于2015年6月发射的哨兵2号（Sentinel－2）多光谱传感器，具有最高可达10 m分辨率、可覆盖13个光谱波段的遥感数据[14]。此外，Sentinel－2数据也是目前唯一一个在“红边”范围含有3个波段的多光谱数据，对监测植被健康信息非常有效。Clevers等[15]对3种多光谱影像的“红边”指数组合进行了比较，基于Sentinel－2的大豆红边叶绿素指数（CIre）与其叶绿素浓度具有极高相关性（R^2 = 0.94）。Verrelst等[16]使用了4种机器学习算法在Sentinel－2的“红边”范围内估算植被叶绿素，R^2 达到了0.99。此外，李粉玲等[17]使用模拟高分1号的遥感光谱指

① 收稿日期：2018－11－27；　修订日期：2019－03－07
基金项目：国家自然科学基金资助项目（41771470）；新疆自治区重点实验室专项基金资助项目（2016D03001）；自治区科技支疆项目（201591101）
作者简介：顾峰（1994－），男，硕士研究生，主要从事干旱区生态环境遥感方面的研究. E-mail：gufeng947210@163.com
通讯作者：丁建丽. E-mail：watarid@xju.edu.cn

数在关中地区建立了冬小麦叶片 SPAD 估算模型，指出基于随机森林算法构建的模型具有最优精度。这些研究成果证明了 Sentinel－2 的空间分辨率和光谱特性在 SPAD 遥感估算上的可行性，与机器学习算法的结合克服了样点数量较少、数据分布不均等不利条件。然而，上述研究多在非干旱地区开展，且大多只利用单一红边参数建模，没有深入考虑土壤属性的多种重要协变量，对干旱区绿洲植被 SPAD 的建模效果有待进一步的考证。

基于此，本研究以渭干河—库车河绿洲(渭—库绿洲)为研究靶区，基于 Sentinel－2 数据和原位 SPAD 数据，设立 3 种组合方案并利用随机森林算法构建了区域棉花、芦苇、大枣、杨树 4 种典型植被叶片的 SPAD 估算模型，以期为干旱区植被叶绿素含量的无损监测以及当地精准农业提供数据支持和技术支撑。

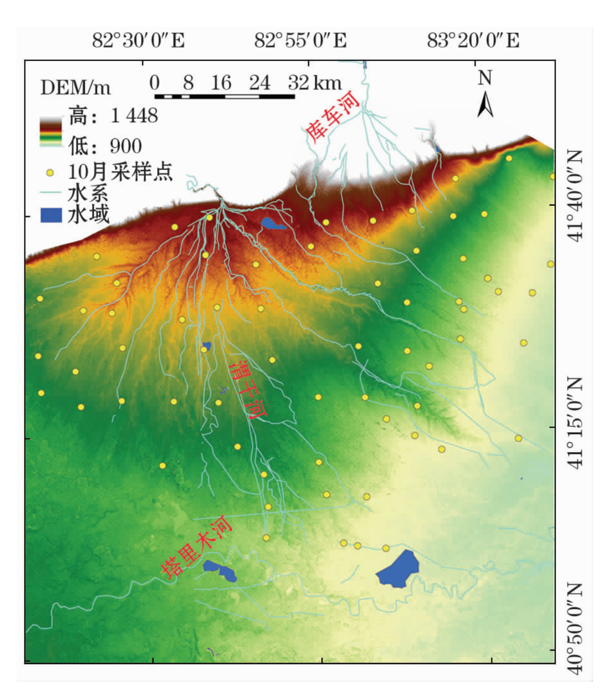
1 数据与方法

1.1 研究区概况

渭—库绿洲(40°50′～41°38′N，81°26′～83°17′E)位于新疆南部阿克苏地区，天山南脉中段塔里木盆地北缘，塔里木河中游(图 1)。绿洲涵盖库车、沙雅、新和 3 县，是典型的洪积－冲积倾斜平原，境内北高南低，平均海拔 971 m。气温日较差大，2001—2016 年间年平均气温 10.3～14.4 ℃，最高气温 41.5 ℃；潜在蒸发量 2 420 mm，多年平均降水量为 43 mm，蒸降比约 54∶1，降水季节性差异很大，多集中于 6—7 月，其他月份降水大幅减少，属于干旱与极端干旱区；土壤主要为潮土、草甸土、风沙土、盐土[18－19]。区域内主要水系为渭干河、库车河、木扎提河、英达雅河及塔里木河，水域多建有平原水库，如库车跃进水库、沙雅齐满水库。绿洲境内为该地主要农作物产区，人工植被以棉花(*Gossypium hirsittum*)、杨树(*Populus* L)、枣树(*Ziziphus jujube*)为主，天然植被以芦苇(*Phragmites australis*)、骆驼刺(*Alhagi sparsifolia*)、花花柴(*Karelina caspica*)、柽柳(*Tamarix chinensis*)、盐爪爪(*Kalidium foliatum*)等耐盐植被为主[20]。

1.2 实测数据

2017 年 10 月 8—19 日，在渭—库绿洲进行了野外实地调查。依据研究区地表形态和地貌差异以及样点分布的均匀性，对绿洲范围内的66个远离其他明显特征(道路，河渠)的点进行采样，坐标通过 GPS 定位并记录。每个采样点测取了叶绿素相对含量(SPAD)、表层土壤有机质(soil organic matter，SOM)、土壤含水量(soil moisture content，SMC)以及土壤电导率(electrical conductivity，EC)等数据。



注:地理底图源自新疆维吾尔自治区测绘地理信息局〔审图号:新 S(2016)118 号〕，未对边界进行修改。

图 1 研究区及采样点

Fig. 1 The study area and sampling sites

1.2.1 植被叶片 SPAD 值 使用日本美能达公司生产的 SPAD 快速叶绿素测定仪，从东、西、南、北方向以及植株冠层的上、中、下，采集 25 片健康、规则的叶片测定 SPAD 值，并将 25 个测定值的算术平均值作为对应采样点特定植被叶片的 SPAD 值。

1.2.2 植被叶片光谱 植被叶片光谱数据由 ASD Field Spec 4 测量便携式光谱仪测定，测定时利用植被探头与叶片夹和光谱仪自带光源，将每个样点(25 片)的每片叶片测定 2 次，注意避开叶脉，保证叶片平整且被探测的面积相同。采集结束后，对光谱数据进行 Savitaky-Golay 滤波，以去除噪声并对光谱曲线进行平滑，然后取均值得到该样点的叶片光谱[21]。

1.2.3 土壤含水量 对每个采样点使用 5 点采样法采集 5 个 15 cm 土层的土样，共采集 330 个土壤样品，记录周边环境信息。在实验室，将土壤样本置于 105 ℃的恒温烘干箱内烘 48 h，然后对同一采样

点土壤水分实测值求平均值，得到该采样点的真实土壤含水量。

1.2.4 土壤有机质 将采集的土样在实验室自然风干、研磨，并剔除土样中的杂质后过 0.25 mm 孔径筛。采用重铬酸钾容量法－外加热法(油浴)测定分选好的土壤有机质含量。

1.2.5 土壤电导率 每组土样称取 3 份样品，每份为 0.5 g，将每份样品置于消煮管中，用去离子水定标至 20 mL，将所有消煮管放置在摇床上，震动处理 1 h，静置 30 min，采用德国 Wissenschaftlich Technische Werkstätten 公司生产的 Cond 7310 电导仪测定，并将 3 份样品的电导率求平均，得到该组土样的电导率[20]。

1.3 影像数据

根据实地采样时间和采样范围，选择 2017 年 10 月 14 日的 2 景 Sentinel－2A 影像(以下简称 S2A)，数据来源于欧洲太空局(http://www.esa.int/ESA)。影像基本无云，质量高。获得的 S2A 多光谱影像数据类型是 Level－1C，是已经过几何校正的反射率数据(表 1)。

表 1 本研究使用的 S2A 数据

Tab. 1 S2A data used in this study

影像编号	获取日期	获取卫星	波段数
N0205_R062_T44TPL_20171014T052639	2017－10－14	S2A	13
N0205_R062_T44TPM_20171014T052639	2017－10－14	S2A	13

1.3.1 图像预处理 使用 ENVI 5.5 软件对影像进行地形校正、辐射定标和 Flaash 大气校正，获得其反射率数据。从表 2 和图 2 可以看出，S2A 波段的空间分辨率不一致，本文使用 3 次卷积内插方法将大气校正后的波段重采样至效果最好的 10 m 空间分辨率。最后，使用 GS(gram-schmidt pan sharpening)融合方法将 2 幅影像拼接，并按照研究区范围进行裁剪。

1.3.2 植被影像反射率的提取和精度 使用大气校正处理后的影像，其植被光谱显示了绿色植被光谱的典型趋势(图 2)。蓝波段(490 nm)和红波段(665 nm)显示出 2 条吸收带，即光辐射光合作用形成 2 个叶绿素吸收层；在绿波段(560 nm)下小的反射峰代表了绿叶植物叶绿素对绿光的强烈反射效应；在 705～783 nm 范围内，“红边”明显并且在近红外波段(865 nm)形成了更高的反射峰。利用大气校正模型减弱了大气对遥感影像的影响，恢复了植被光谱曲线的典型趋势。

表 2 S2A 数据信息

Tab. 2 Spectral bands and resolutions of S2A sensors

波段	S2A 中心波长/nm	S2A 波段宽度/nm	空间分辨率/m
海岸波段(*B1*)	443.9	27	60
蓝波段(*B2*)	496.9	98	10
绿波段(*B3*)	560.0	45	10
红波段(*B4*)	664.5	38	10
植被红边 1 波段(*B5*)	703.9	19	20
植被红边 2 波段(*B6*)	740.2	18	20
植被红边 3 波段(*B7*)	782.5	28	20
近红外波段(*B8*)	835.1	145	10
近红外波段(*B8A*)	864.8	33	20
水汽波段(*B9*)	945.0	26	60
卷云波段(*B10*)	1 373.5	75	60
短波红外 1(*B11*)	1 613.7	143	20
短波红外 2(*B12*)	2 202.4	242	20

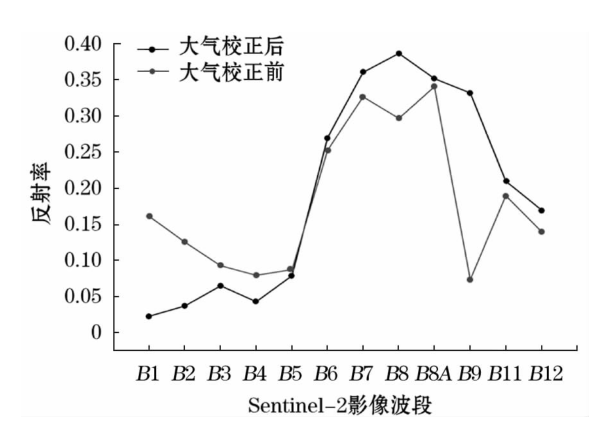


图 2 大气校正前后的植被像元反射率

Fig. 2 Reflectance of vegetation pixels before and after correction

根据现场测量的 GPS 坐标，将位于 S2A 中的采样点控制在一个像元中。采样点的 GPS 坐标位于两个或多个像素的边界处，将相邻像元的光谱反射率的平均值作为样本点的影像反射率；坐标位于像元的中心位置，将像元的光谱反射率认定为样本点的影像反射率。如图 3 所示，S2A 每组波段光谱响应效果较好的位置并不在波段中心波长位置处，因此要将样本点的光谱仪实测反射率重采样，与 S2A 各波段光谱响应较好的波长位置匹配，然后根据该点的影像反射率，按如下公式求出建立植被指数所需各波段的反射率相对误差：

$$E_i = \left| \frac{R_{ia} - R_{ib}}{R_{ia}} \right| \times 100\% \tag{1}$$

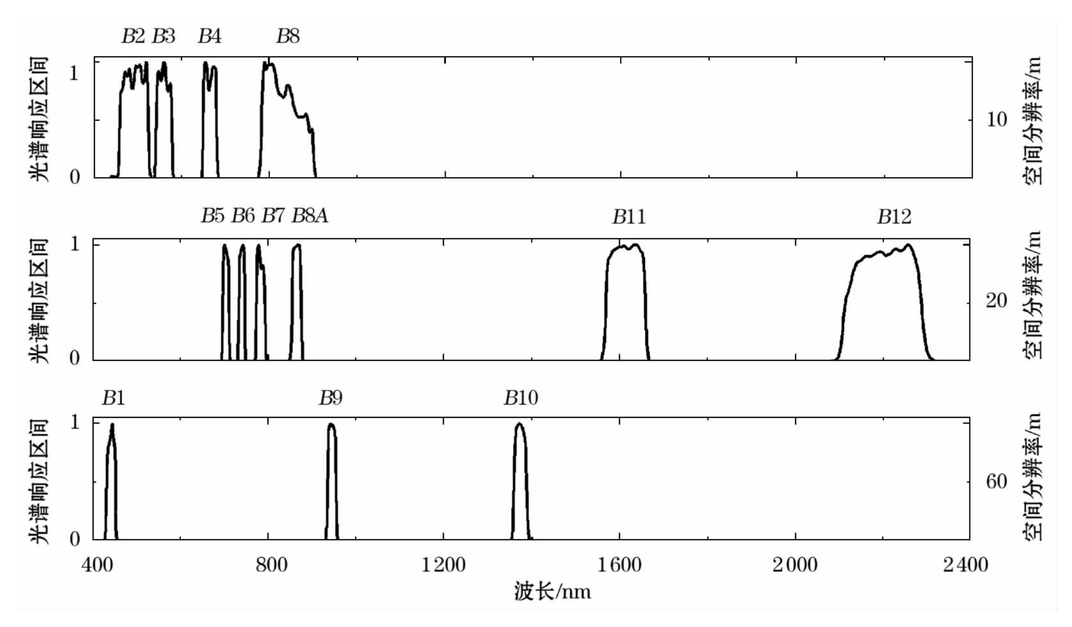


图 3　S2A 影像的波谱响应关系

Fig. 3　Spectral response relationship of S2A images

式中:E_i 为第 i 波段的相对误差;R_{ia}是第 i 波段对应的实测光谱反射率;R_{ib}是第 i 波段对应的大气校正后的光谱反射率[22]。结果表明,图像处理可以获得接近实测的叶片光谱(表 3),这为后续分析提供了保证。

表 3　植被叶片光谱反射率与实测反射率间的相对误差

Tab. 3　Relative errors between vegetation canopy spectral reflectance and measured reflectance　/%

植被	B2	B3	B4	B5
芦苇	14.07	36.47	2.07	3.86
棉花	17.31	22.97	16.14	11.25
大枣	12.48	28.62	7.83	10.51
杨树	19.77	27.83	35.14	7.63
植被	B6	B7	B8	B8A
芦苇	10.86	2.14	2.83	10.29
棉花	9.14	6.77	8.52	9.37
大枣	8.32	6.98	5.24	8.73
杨树	9.97	12.70	15.40	10.50

1.3.3　构建和选择植被指数　植被指数是将遥感影像上不同波段的光谱反射率使用线性或非线性组合,得到反映植被某种特征信息的指数。植被叶片光谱曲线在 “红边”范围出现的急速上升的陡坡,是一阶导数光谱曲线在该区间内的拐点[23]。考虑到植被光谱特性和 S2A 影像的“红边”优势,建立 23 个植被指数[24－29],并对研究区的 S2A 影像进行一阶、二阶求导(图 4),再分别建立相应的 23 种指数关系(表 4)。

1.4　构建估算模型

分别建立对采样点棉花、芦苇、杨树、大枣叶片的 SPAD 估算模型,从现场随机抽取的总样本中,随机选取总数 75% 的样本建立估算模型。其他 25% 的样本用于模型试验(表 5)。

随机森林算法(random forest, RF)作为一种较新的集成学习算法,可以综合利用多种特征变量建模并评价。将野外测得的植被冠层 SPAD 值作为因变量;植被指数系列 1(原影像的 23 个植被指数),植被指数系列 2(原一阶微分影像、二阶微分影像的 23 个植被指数),土壤参数(SOM、SMC、EC)分别作为自变量。设立 3 种方案(表 6),在 matlab 中利用随机森林回归算法分别对采样点的 4 种植被(棉花、芦苇、大枣、杨树)建立植被叶片 SPAD 的估算模型。

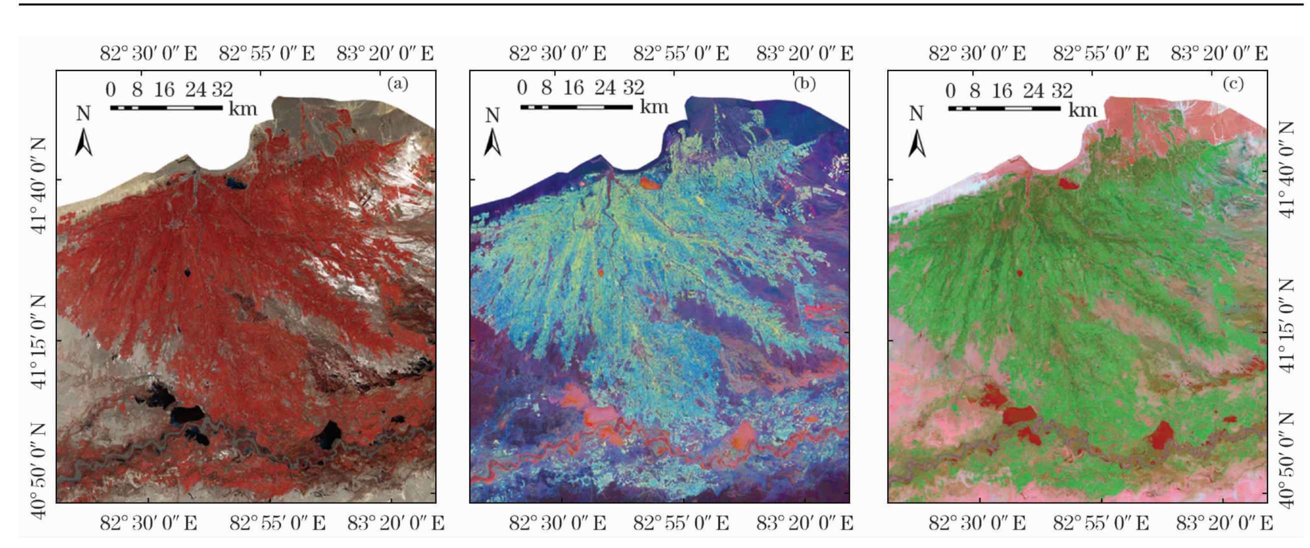
为了消除不同植被指数、土壤参量之间的量纲影响,对数值进行归一化处理,使各数值处于同一数量级,以便进行综合对比评价[30]。归一化函数为:

$$a_n = \frac{a - \min_a}{\max_a - \min_a} \tag{2}$$

式中:a 为需要进行归一化处理的数据;a_n 为归一化后的样本数据;$\max_a$ 为 a 中的最大值;$\min_a$ 为 a 中的最小值。归一化后的数值范围是[0,1]。

1.5　特征重要性评分

在随机森林模型中[29],通过选择原始数据集和训练数据集(即校准数据集)的随机样本,使用确定



注：(a)为假彩色($B8A$,$B4$,$B3$)合成的 S2A；(b)为一阶微分影像同假彩色波段组合一致的 S2A；(c)为二阶微分影像同假彩色波段组合一致的 S2A。

图 4　不同预处理后的 S2A 数据

Fig. 4　S2A data after different pretreatments

表 4　植被指数

Tab. 4　Vegetation indexes

分类	指数	全称	计算公式
传统指数[24]	NDVI	归一化植被指数	$(B8A-B4)/(B8A+B4)$
	RVI	比值植被指数	$B8A/B4$
	DVI	差值植被指数	$B8A-B4$
	CI_g	叶绿素指数－绿色	$B8A/B3-1$
	CVI	叶绿素植被指数	$B8A\times B4/B3^2$
	MTVI2	改良三角植被指数 2	$\frac{1.5[1.2(B8A-B4)-2.5(B4-B3)]}{\sqrt{(2B8A+1)^2-(6B8A-5\sqrt{B4})-0.5}}$
红边指数[25-29]	NDVIre1	红边归一化植被指数 1	$(B8A-B5)/(B8A+B5)$
	NDVIre2	红边归一化植被指数 2	$(B8A-B6)/(B8A+B6)$
	NDVIre3	红边归一化植被指数 3	$(B8A-B7)/(B8A+B7)$
	NDre1	红边归一化差值 1	$(B6-B5)/(B6+B5)$
	NDre2	红边归一化差值 2	$(B7-B5)/(B7+B5)$
	TCIre1	红边三角叶绿素指数 1	$1.2(B5-B3)-1.5(B4-B3)\sqrt{\frac{B5}{B4}}$
	TCIre2	红边三角叶绿素指数 2	$1.2(B6-B3)-1.5(B4-B3)\sqrt{\frac{B6}{B4}}$
	TCIre3	红边三角叶绿素指数 3	$1.2(B7-B3)-1.5(B4-B3)\sqrt{\frac{B7}{B4}}$
	TCARIre1	红边转化叶绿素吸收反射率指数 1	$3[(B5-B4)-0.2(B5-B3)B5/B3]$
	TCARIre2	红边转化叶绿素吸收反射率指数 2	$3[(B6-B4)-0.2(B6-B3)B6/B3]$
	TCARIre3	红边转化叶绿素吸收反射率指数 3	$3[(B7-B4)-0.2(B7-B3)B7/B3]$
	MSRre1	修正的简单比率 1	$\frac{B8A/B5-1}{\sqrt{(\frac{B8A}{B5}+1)}}$
	MSRre2	修正的简单比率 2	$\frac{B8A/B6-1}{\sqrt{(\frac{B8A}{B6}+1)}}$
	MSRre3	修正的简单比率 3	$\frac{B8A/B7-1}{\sqrt{(\frac{B8A}{B7}+1)}}$
	CIre1	红边叶绿素指数 1	$B8A/B5-1$
	CIre2	红边叶绿素指数 2	$B8A/B6-1$
	CIre3	红边叶绿素指数 3	$B8A/B7-1$

表 5　建模方案

Tab. 5　Modeling plans

植被类型	样本总数	训练样本数	估算样本数
棉花	40	30	10
芦苇	45	34	11
杨树	23	17	6
大枣	30	23	7

表 6　实验方案信息

Tab. 6　Information of experimental programs

实验方案	特征组合
1	植被指数系列 1
2	植被指数系列 1 + 植被指数系列 2
3	植被指数系列 1 + 植被指数系列 2 + 土壤参数

性算法构建每颗树。需要在随机森林模型中确定 3 个参数:N 是从原始数据集中抽取的 N 个集合用于生成回归树的数量(默认值为 500 棵树);K 是从 N 个训练集中提取的 K 个特征变量(默认值是总数据量的 66.67%);X 是从 K 个总特征变量中选取的最具重要性的 X 个特征参数作为每棵树的分类节点($X \leqslant K$)。此外,在抽取训练集 X 过程中,未被抽取部分称为袋外数据(out-of-bag,OOB),使用此部分数据计算内部误差(OOB 误差),OOB 误差越小,说明随机森林模型的分类精度越高。OOB 误差也可以计算特征变量的重要性,将需要计算的一个特征变量改变并保持的其他特征变量不变,计算被改变特征变量前、后 OOB 误差的差值和百分比,即为判断该特征变量重要性的依据[27]。其评估模型如下:

$$VI(M_A) = \frac{1}{N}\sum_{t=1}^{N}\left(B_{n_t}^{M_A} - B_{o_t}^{M_A}\right)$$

式中:VI 表示特征变量的重要性;M 为样本的全部特征数;N 为生成的决策树的棵数;$B_{o_t}^{M_A}$ 为任意特征值 M_A 未加入噪声干扰时第 t 棵决策树的 OOB 误差;$B_{n_t}^{M_A}$ 为任意特征值 M_A 加入噪声干扰时第 t 棵决策树的 OOB 误差。特征变量的重要性 VI 值越大,则说明特征 M_A 对分类结果的影响越大。

1.6　模型精度检验方法

本研究利用估算样本对基于不同输入变量的模型精度进行检验,并使用以下指标评估模型效果[31]:确定系数(R^2)、均方根误差($RMSE$)、相对百分比偏差(RPD)。本研究中,R^2 包括建模集的确定系数(R^2cal)和验证集的确定系数(R^2val)。$RMSE$ 包括建模集的均方根误差($RMSE_C$)和验证集的均方根误差($RMSE_V$)。采用 3 类标准评估模型的可预测性,即Ⅰ类($RPD > 2.2$)具有良好的可预测性;Ⅱ类($2.0 < RPD < 2.2$)具有中度可预测性;Ⅲ类($RPD < 2.0$)具有较差的可预测性[32]。

2　结果与分析

2.1　特征重要性分析

在本研究中,当树的数量(N)为 500 时,袋外误差 OOB 基本收紧并趋于稳定,因此,所有实验方案的 N 为 500,K 为 335,X 则设为 2。上文中的方案 1 和方案 2 均为对照实验,不需要做特征变量的重要性评价,故在方案 3 的基础上,对参与建模的 72 个特征变量进行特征重要性评价,并选取前 31 个得分较高的重要性生成重要性分布图(图 5)。

在 4 种植被叶片的 SPAD 估算模型中,棉花、芦苇、枣树的一阶指数的重要性都是各自建模特征中占比最大的,尤其是芦苇,达到了 42%;而土壤参量虽然只有 3 个特征变量(SMC、SOM、EC)参与,占所有特征变量总数的 4.17%,在每种估算模型中贡献的重要性都高于 6%,特别是杨树叶片的 SPAD 估算模型,其占比达到了 15%,且 SMC 和 EC 是重要性前 2 位的特征变量,分别到达了 5.43% 和 5.01%(图 6)。可见,土壤参量对植被叶片的叶绿素相对含量有较强的响应。

2.3　叶片 SPAD 值模型估算结果

在所有植被叶片 SPAD 估算模型中,从方案 1 到方案 3 的建模集效果逐步提升,R^2_{CAL} 逐渐增大,$RMSE_C$ 也呈现出逐渐变小的趋势(图 7)。加入微分影像的植被指数和实测土壤属性后,建模效果显著提高。在预测效果方面,所有估算模型的 RPD 都达到了Ⅱ类标准,具有中度及以上的可预测性,但与建模集效果都逐渐提高的情况不同,芦苇、枣树、杨树叶片 SPAD 估算模型的验证集中,相比方案 1,方案 2 的 RPD 反而减小,其可预测性出现了下降的情况,尤其是芦苇叶片 SPAD 估测模型的 RPD 值,由方案 1 的 2.620 骤降到方案 2 的 2.009,模型的可预测性也从良好变成了中等,可能是方案 2 加入的植被指数较多,导致了部分特征变量冗余,降低了模型的预测效果(表 7)。此外,杨树叶片 SPAD 估算模型的方案 2,R^2 仅比方案 1 提升了 0.01,但其 $RMSE_V$ 却增大了 0.724,RPD 下降了 0.014,可见其方案 2 的模型整体效果较方案 1 下降。而所有植被在方案 3 的验证效果却都是 3 种方案中最优的,R^2 最

重要性得分

0 2 4 6 8

特征变量

(a) 枣树

1st CIg
SMC
2nd_TCARIre1
2nd CIre3
2nd_MSRre3
2nd CIre1
1st DVI
1st_TCARIre3
1st_TCARIre2
1st_TCIre3
1st_CIre1
2nd_MSRre1
2nd_TCARIre2
TCARIre3
1st_TCARIre1
1st_TCIre3
1st_NDVIre3
CIre2
TCARIre2
NDVIre2
TCIre1
TCARIre1
EC
2nd_NDVI
2nd_MTVI2
2nd_TCIre1
2nd_DVI
DVI
1st_CIre3
NDVIre1
MTVI2

重要性得分

0 2 4 6

特征变量

(b) 芦苇

2nd_NDre1
1st_DVI
1st_NDVI
1st_TCARIre2
SMC
1st_TCARIre1
1st_NDre2
SOM
1st_TCIre2
1st_CIre2
2nd_CIre1
2nd_MTVI2
1st_NDVIre2
1st_NDre1
1st_NDVIre3
2nd_MSRre2
1st_TCIre3
EC
1st_CIre1
2nd_CLg
TCARIre3
2nd_TCARIre1
NDre2
2nd_TCIre1
NDVI
2nd_NDVIre1
DVI
2nd_NDre2
TCARIre2
1st_NDVIre1
2nd_CVI

重要性得分

0 2 4 6 8

特征变量

(c) 棉花

1st_NDVI
2nd_MTVI2
SMC
SOM
CIre3
1st_NDVIre3
MSRre3
1st_NDVIre1
1st_NDVIre2
NDVIre2
1st_CIre1
NDVIre3
RVI
CIre2
1st_TCARIre3
TCIre2
2nd_DVI
TCARIre2
1st_NDre2
MSRre2
DVI
2nd_NDre1
1st_NDre1
2nd_RVI
1st_TCIre3
2nd_NDVI
1st_CIre2
1st_TCARIre1
MSRre1
EC
2nd_TCARIre1

重要性得分

0 2 4 6

特征变量

(d) 杨树

SMC
EC
1st_NDVIre3
TCIre1
SOM
2nd_MSRre3
2nd_CIre3
1st_TCARIre3
2nd_CIre2
2nd_TCIre1
CARIre1
MSRre3
2nd_NDVIre2
2nd_TCIre3
2nd_NDVIre1
1st_NDre1
2nd_MSRre2
1st_TCIre3
CIre3
2nd_CIre1
NDVIre3
1st_NDVI
TCARIre2
2nd_NDVI
2nd_TCARIre1
2nd_MTVI2
MSRre2
DVI
2nd_CLg
TCIre3
NDre1

注:1st 表示已经过一阶导之后的影像,2nd 表示经过二阶导之后的影像。

图 5 特征重要性分布

Fig. 5 Distribution of characteristic importance

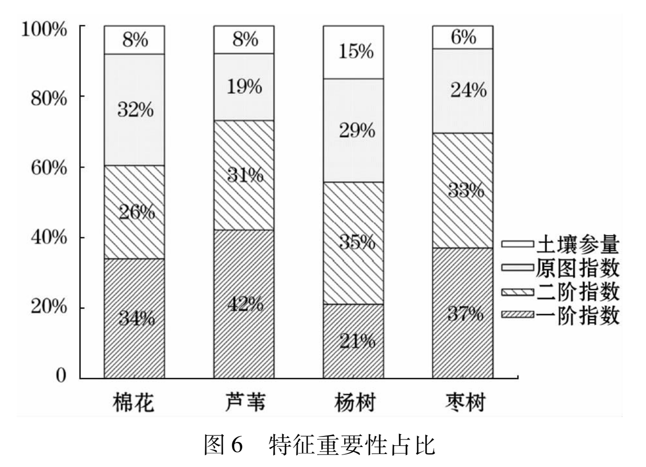
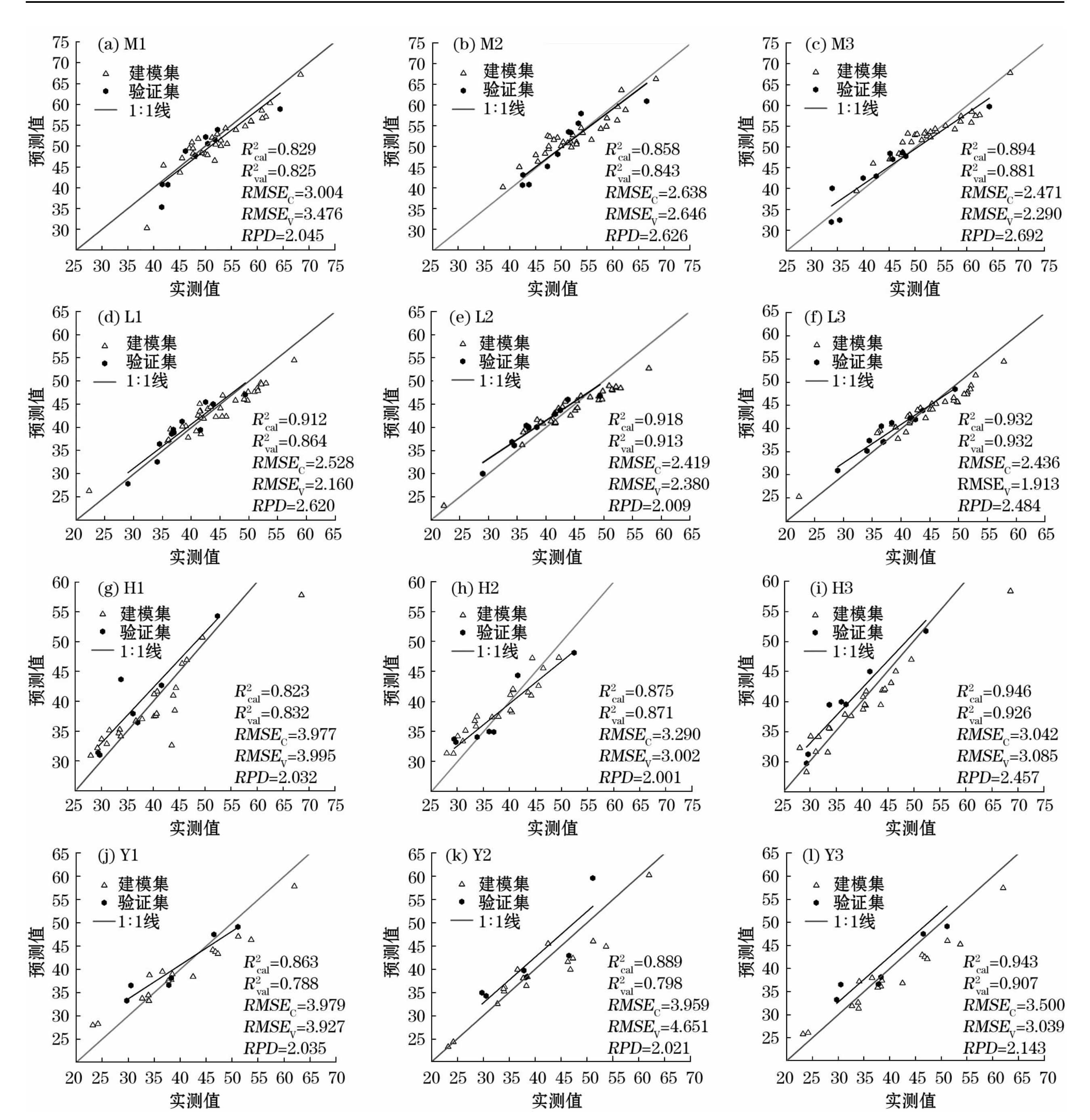


图 6 特征重要性占比

Fig. 6 Proportions of characteristic importance

大,$RMSE_V$ 最小,$RPD > 2.2$。这说明方案 3 加入的土壤参量不是冗余变量,且对估算叶片 SPAD 值具有积极影响。

图 8 为利用方案 3 的 4 种植被叶片 SPAD 值预测模型制作的研究区植被叶片 SPAD 遥感监测专题图。以同步采集的地面实测数据进行精度检验:将地面实测 SPAD 值与叶绿素含量分布图上同样点的估算值进行回归拟合。分布图估算值与实测数据的拟合方程决定系数 R^2 为 0.826,$RMSE$ 为 3.794,结果表明,基于估算模型的植被叶片 SPAD 值空间分布与实测值基本一致。



注：M 代表棉花叶片，L 代表芦苇叶片，H 代表枣树叶片，Y 代表杨树叶。1 代表方案 1，2 代表方案 2，3 代表方案 3，M1 是方案 1 的棉花叶片 SPAD 估算模型，以此类推。R^2cal 表示建模集的确定系数，R^2val 表示验证集的确定系数，$RMSE_C$ 表示建模集的均方根误差，$RMSE_V$ 表示验证集的均方根误差，RPD 是验证集的相对百分比偏差，SPAD 表示叶绿素相对含量。

图 7　叶片 SPAD 预测值与实测值分布

Fig. 7　Distribution of estimated and measured leaf SPAD values

3　讨 论

SPAD 值反映了植被叶片单位表面积的叶绿素含量，并可以表征植被叶片、冠层的养分和长势状况，它与植被叶片的反射光谱之间存在较强的相关性[6]，选取特征波段构建光谱指数在不同程度上降低了由环境变化而引起的反射率误差，可以用来定量反演植被的理化参数。

（1）结合图 1 和图 8，绿洲 SPAD 值的空间分布大体上为内部高，四周低，且受水系的影响较大。河流水系及水域周边植被的叶绿素含量明显较高，在 50～60 之间，而研究区西北角处叶绿素较高且没有主要水系，可能是由于暂时性灌溉导致的。要对 SPAD 在空间尺度上进行更深入分析，除了掌握主

表 7 叶片 SPAD 预测值与实测值分布

Tab. 7 Distribution of estimated and measured leaf SPAD values

植被类型	方案	建模集		验证集		
		R^2_{CAL}	$RMSE_C$	R^2_{VAL}	$RMSE_V$	*RPD*
棉花	方案 1	0.829	3.004	0.825	3.476	2.045
	方案 2	0.858	2.638	0.843	2.646	2.626
	方案 3	0.894	2.471	0.881	2.290	2.692
芦苇	方案 1	0.912	2.528	0.864	2.159	2.620
	方案 2	0.918	2.419	0.913	2.380	2.009
	方案 3	0.932	2.436	0.932	1.913	2.484
枣树	方案 1	0.823	3.977	0.832	3.995	2.032
	方案 2	0.875	3.290	0.871	3.002	2.001
	方案 3	0.946	3.042	0.926	3.085	2.457
杨树	方案 1	0.863	3.979	0.788	3.927	2.035
	方案 2	0.889	3.959	0.798	4.651	2.021
	方案 3	0.943	3.500	0.907	3.039	2.143

注：R^2_{cal} 表示建模集的确定系数；R^2_{val} 表示验证集的确定系数；$RMSE_C$ 表示建模集的均方根误差；$RMSE_V$ 表示验证集的均方根误差；*RPD* 是验证集的相对百分比偏差。

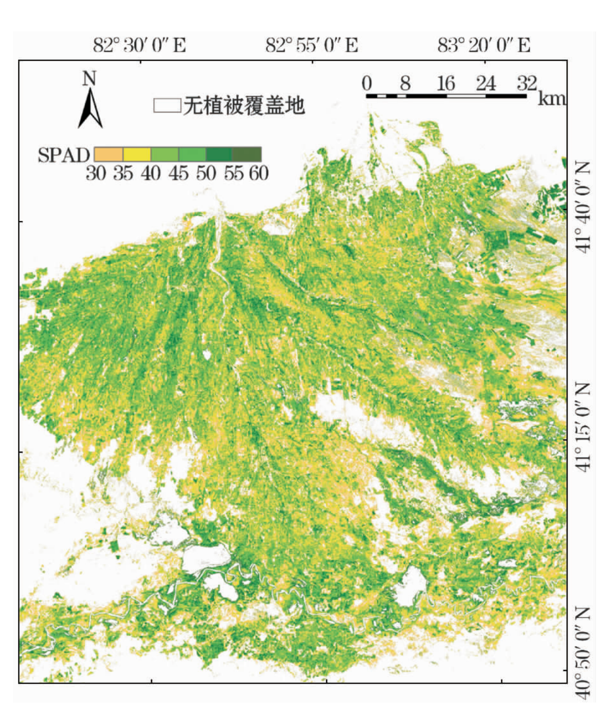


图 8 叶片 SPAD 遥感监测专题图

Fig. 8 Spatial distribution of leaf SPAD values at greenup stage

要自然水系，还应掌握该地的灌溉区划。

（2）利用多光谱数据对叶绿素的遥感估算，以往研究选取的光谱指数类型较少，对叶绿素敏感的红边光谱指数更少[33]。本文选取 23 种光谱指数参与模型构建，发现基于多个植被指数的 SPAD 估算模型效果（$R^2 \geq 0.8$）明显好于以往基于单个植被指数的 SPAD 估算模型。且基于“红边”的光谱指数对估算模型起到了更为关键的作用，更能反映叶绿素较高时的吸收差异。而在一阶微分影像上建立的传统光谱指数（1st_DVI，1st_NDVI，1st_RVI），在模型中的重要性比在原影像上（DVI，NDVI，RVI）有了大幅提升。这是由于植被反射光谱在“红边”处的斜率变化剧烈，一阶微分影像的光谱指数反映出了“陡坡”上下的斜率差异。本文使用的指数是在前人的植被指数模型基础上结合 S2A 构建的，并未深入研究光谱指数的机理，致使部分重要性偏低的指数也参与了模型构建，如芦苇叶片在绿波段（*B3*）与实测光谱相对误差达到了 36.47%，绿波段（*B3*）参与构建的植被指数导致模型的预测能力偏低，今后需在优选特征的基础上进行模型构建。

（3）除光谱指数外，土壤属性也对预测结果有较大影响。在估算模型加入土壤含水量（SMC）、土壤有机质（SOM）、土壤电导率（EC）3 个土壤参量后，可预测性显著提高，特别是 SMC 在 4 种植被叶片 SPAD 的估算模型中，重要性都在前 5 位，可见，将易获取的土壤参量加入模型能提升估算精度。

（4）本文采用的随机森林回归算法是估算植被生物量的优选方法[34]。它能最大程度地综合多种特征变量，因此影像数据覆盖的范围越广、空间分辨率和光谱分辨率越高，模型效果越好。且随机森林算法能给出不同变量对模型的影响评分，这替代并且升华了以往线性模型的相关性，还可将模型优选出的重要性排序靠前的特征变量与估算对象做更深入的研究。构建的模型实现了一定范围、特定时间上的 SPAD 值的快速估算，但实现更大范围更高精度的 SPAD 值估算还需要利用多年的新数据加以改进。

4 结 论

在 12 个基于随机森林回归算法的植被叶片叶绿素相对含量（SPAD）估算模型中，所有红边指数和土壤属性参数的组合展现出最佳 SPAD 预测能力，芦苇叶片表现效果最佳（$R^2_{CAL} = 0.932$，$RMSE_C = 2.436$，$R^2_{VAL} = 0.932$，$RMSE_V = 1.913$，$RPD = 2.484$）。而一阶微分指数和土壤含水量（SMC）在随机森林重要性排序中，与 SPAD 具有显著非线性关系。依据该模型制作的研究区植被叶片 SPAD 遥感监测专题图，实现了植被叶片 SPAD 值在空间尺度上的表达，为多光谱遥感监测植被叶绿素含量提供了参考。综上所述，S2A 数据具有较强的预测绿洲植被 SPAD 值的潜力，基于多种光谱指数和土壤

参量组合的随机森林模型为监测、估算和预测干旱和半干旱地区绿洲中植被的 SPAD 提供了一种快速,经济且稳定的方法。

参考文献(References):

［1］ Hanelt D. Photosynthesis assessed by chlorophyll fluorescence［J］. Bioassays,2018,119:169－198.

［2］ Jhanji S,Sekhon N K. Evaluation of potential of portable chlorophyll meter to quantify chlorophyll and nitrogen contents in leaves of wheat under different field conditions［J］. Indian Journal of Experimental Biology,2018,56(10):750－758.

［3］ Chang S X,Robison D J. Nondestructive and rapid estimation of hardwood foliar nitrogen status using the SPAD-502 chlorophyll meter［J］. Forest Ecology and Management,2003,181(3):331－338.

［4］ Songsri P,Jogloy S,Holbrook C C,et al. Association of root,specific leaf area and SPAD chlorophyll meter reading to water use efficiency of peanut under different available soil water［J］. Agricultural Water Management,2009,96(5):790－798.

［5］ Swiader J M,Moore A. SPAD-chlorophyll response to nitrogen fertilization and evaluation of nitrogen status in dryland and irrigated pumpkins［J］. Journal of Plant Nutrition,2002,25(5):1 089－1 100.

［6］ Yuan S,Goron T L,Huang L,et al. Rice leaf lateral asymmetry in the relationship between SPAD and area-based nitrogen concentration［J］. Symmetry,2017,9(6):83.

［7］ 王娟,韩登武,任岗,等. SPAD 值与棉花叶绿素和含氮量关系的研究［J］. 新疆农业科学,2006,43(3):167－170.［Wang Juan,Han Dengwu,Ren Gang,et al. A study on relation between SPAD value,chlorophyll and nitrogen content in cotton［J］. Xinjiang Agricultural Sciences,2006,43(3):167－170.］

［8］ 邓兴耀,姚俊强,刘志辉. 基于 GIMMS NDVI 的中亚干旱区植被覆盖时空变化［J］. 干旱区研究,2017,34(1):10－19.［Deng Xingyao,Yao Junqiang,Liu Zhihui. Spatiotemporal dynamic change of vegetation coverage in arid regions in central Asia based on GIMMS NDVI［J］. Arid Zone Research,2017,34(1):10－19.］

［9］ 邹红玉,郑红平. 浅述植被"红边"效应及其定量分析方法［J］. 遥感信息,2010(4):112－116.［Zou Hongyu,Zheng Hongping. The effect and method of quantitative analysis of "Red Edge" of vegetation［J］. Remote Sensing Information,2010(4):112－116.］

［10］ 程乾,黄敬峰,王人潮,等. MODIS 植被指数与水稻叶面积指数及叶片叶绿素含量相关性研究［J］. 应用生态学报,2004,15(8):1 363－1 367.［Chen Qian,Huang Jingfeng,Wang Renchao,et al. Correlation analysis of simulated MODIS vegetation indices and rice leaf area index and leaf chlorophyll content［J］. Chinese Journal of Applied Ecology,2004,15(8):1 363－1 367.］

［11］ Shou L N,Jia L L,Chen X P,et al. Using high-resolution satellite image to evaluate nitrogen status of winter wheat in the North China plain［J］. Journal of Plant Nutrition,2007,30(10):1 669－1 680.

［12］ 宋晓宇,黄文江,王纪华,等. ASTER 卫星遥感影像在冬小麦品质监测方面的初步应用［J］. 农业工程学报,2006,22(9):148－153.［Song Xiaoyu,Huang Wenjiang,Wang Jihua,et al. Preliminary application of ASTER images in winter wheat quality monitoring［J］. Transactions of the CSAE,2006,22(9):148－153.］

［13］ 谭昌伟,王纪华,赵春江,等. 利用 Landsat TM 遥感数据监测冬小麦开花期主要长势参数［J］. 农业工程学报,2011,27(5):224－230.［Tan Changwei,Wang Jihua,Zhao Chunjiang,et al. Monitoring wheat main growth parameters at anthesis stage by Landsat TM［J］. Transactions of the CSAE,2011,27(5):224－230.］

［14］ Delegido J,Verrelst J,Alonso L,et al. Evaluation of sentinel22 red-edge bands for empirical estimation of green LAI and chlorophyll content［J］. Sensors,2011,11(7):7 063－7 081.

［15］ Clevers J G P W,Gitelson A A. Remote estimation of crop and grass chlorophy Ⅱ and nitrogen content using red-edge bands on Sentinel－2 and －3［J］. International Journal of Applied Earth Observation and Geoinformation,2013,23:344－351.

［16］ Verrelst J,Muñoz J,Alonso L,et al. Machine learning regression algorithms for biophysical parameter retrieval:Opportunities for Sentinel－2 and－3［J］. Remote Sensing of Environment,2012,118:127－139.

［17］ 李粉玲,王力,刘京,等. 基于高分一号卫星数据的冬小麦叶片 SPAD 值遥感估算［J］. 农业机械学报,2015,46(9):273－281.［Li Fenling,Wang Li,Liu Jing,et al. Remote sensing estimation of SPAD value for wheat leaf based on GF－1 data［J］. Transactions of the Chinese Society for Agricultural Machinery,2015,46(9):273－281.］

［18］ 王丹丹,程猛,杨瑞红,等. 近 20 a 渭干河绿洲土壤盐分变化特征［J］. 干旱区研究,2015,32(6):1 076－1 081.［Wang Dandan,Cheng Meng,Yang Ruihong,et al. Changing characteristics of soil salt in Weigan river oasis for the last 20 years［J］. Arid Zone Research,2015,32(6):1 076－1 081.］

［19］ 康璇,王雪梅,柴仲平. 近 25 a 来渭—库绿洲土地利用/覆被变化及其影响因素［J］. 水土保持通报,2016,36(5):333－339.［Kang Xuan,Wang Xuemei,Chai Zhongping. Land use/cover changes and Influencing factors in delta oasis of Weigan-Kuqa river during last 25 years［J］. Bulletin of Soil and Water Conservation,2016,36(5):333－339.］

［20］ 丁建丽,瞿娟,孙永猛,等. 基于 MSAVI-WI 特征空间的新疆渭干河—库车河流域绿洲土壤盐渍化研究［J］. 地理研究,2013,32(2):223－232.［Ding Jianli,Qu Juan,Sun Yongmeng,et al. The retrieval model of soil salinization information in arid region based on MSAVI-WI feature space:A case study of the delta oasis in Weigan-Kuqa watershed［J］. Geographical Research,2013,32(2):223－232.］

［21］ 杨爱霞,丁建丽. 新疆艾比湖湿地土壤有机碳含量的光谱测定方法对比［J］. 农业工程学报,2015,31(18):162－168.［Yang Aixia,Ding Jianli. Comparative assessment of two methods for estimation of soil organic carbon content by Vis-NIR spectra in Xinjiang Ebinur Lake Wetland［J］. Transactions of the Chinese Society of Agricultural Engineering,2015,31(18):162－168.］

［22］ Green M. A relative error bound for balanced stochastic truncation［J］. IEEE Transactions on Automatic Control,1988,33(10):961－965.

［23］ Filella I,Penuelas J. The red edge position and shape as indicators of plant chlorophyll content,biomass and hydric status［J］. International Journal of Remote Sensing,1994,15(7):1 459－1 470.

［24］ Shang J,Liu J,Ma B,et al. Mapping spatial variability of crop growth conditions using RapidEye data in Northern Ontario,Canada［J］. Remote Sensing of Environment,2015,168:113－125.

［25］ Gitelson A A,Gritz Y,Merzlyak M N. Relationships between leaf

chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves[J]. Journal of Plant Physiology,2003,160(3):271 – 282.

[26] Jiang Z,Huete A R,Didan K,et al. Development of a two-band enhanced vegetation index without a blue band[J]. Remote Sensing of Environment,2008,112(10):3 833 – 3 845.

[27] Haboudane D,Tremblay N,Miller J R,et al. Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data[J]. IEEE Transactions on Geoscience and Remote Sensing,2008,46(2):423 – 437.

[28] Vincini M,Frazzi E,D'Alessio P. A broad-band leaf chlorophyll vegetation index at the canopy scale[J]. Precision Agriculture,2008,9(5):303 – 319.

[29] Wu C,Niu Z,Tang Q,et al. Estimating chlorophyll content from hyperspectral vegetation indices:Modeling and validation[J]. Agricultural and Forest Meteorology,2008,148(8 – 9):1 230 – 1 241.

[30] 程志庆,张劲松,孟平,等. 杨树叶片叶绿素含量高光谱估算模型研究[J]. 农业机械学报,2015,46(8):264 – 271. [Cheng Zhiqing,Zhang Jinsong,Meng Ping,et al. Hyperspectral estimation model of chlorophyll content in poplar leaves[J]. Transactions of the Chinese Society for Agricultural Machinery,2015,46(8):264 – 271.]

[31] 葛翔宇,丁建丽,王敬哲,等. 基于竞争适应重加权采样算法耦合机器学习的土壤含水量估算[J]. 光学学报,2018,38(10):1 – 8. [Ge Xiangyu,Ding Jianli,Wang Jingzhe,et al. Estimation of soil moisture content based on competitive adaptive reweighted sampling algorithm coupled with machine learning[J]. Acta Optica Sinica,2018,38(10):1 – 8.]

[32] Armstrong J S,Collopy F. Error measures for generalizing about forecasting methods:Empirical comparisons[J]. International Journal of Forecasting,1992,8(1):69 – 80.

[33] Main R,Cho M A,Mathieu R,et al. An investigation into robust spectral indices for leaf chlorophyll estimation[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2011,66(6):751 – 761.

[34] Mutanga O,Adam E,Cho M A. High density biomass estimation for wetland vegetation using WorldView – 2 imagery and random forest regression algorithm[J]. International Journal of Applied Earth Observation and Geoinformation,2012,18:399 – 406.

Estimation of Chlorophyll Content of Typical Oasis Vegetation in Arid Area Based on Sentinel – 2 Data

GU Feng[1,2,3], DING Jian-li[1,2,3], GE Xiang-yu[1,2,3], GAO Shi-bao[4], WANG Jing-zhe[1,2,3]

(1. *College of Resources and Environment Sciences, Xinjiang University, Urumqi* 830046, *Xinjiang, China*; 2. *Key Laboratory of Wisdom City and Environmental Modeling under Department of Education of Xinjiang Uygur Autonomous Region, Urumqi* 830046, *Xinjiang, China*; 3. *Key Laboratory of Oasis Ecology under the Ministry of Education, Xinjiang University, Urumqi* 830046, *Xinjiang, China*; 4. *School of Earth Sciences, Zhejiang University, Hangzhou* 310027, *Zhejiang, China*)

Abstract: The Ogan-Kuqa River Delta Oasis, a typical oasis in the arid zone in China, was taken as the study area. The method of Random Forest with a comparative advantage in machine learning was chosen to model and estimate the relative contents of chlorophyll (SPAD values) of leaves from four kinds of representative vegetation (cotton, reed, poplar and jujube). The 23 broadband spectral indices of vegetation, which are sensitive to chlorophyll content, were obtained based on the reflectance of original Sentinel – 2 image with rich spectral information in the "red edge" bands. These vegetation indices were extracted again in the original band order on the firstly-derived Sentinel – 2 image and secondly-derived Sentinel – 2 image. Three soil parameters (soil moisture content, SMC; soil organic matter, SOM; electrical conductivity, EC) related to vegetation growth were all conducted as the characteristic variables affecting SPAD values. According to the characteristic variables above, three modelling schemes could be developed to monitor the SPAD values of vegetation leaves in oasis. The results showed that: ① Vegetation indices obtained from the firstly-derived image played a more important role than the original vegetation indices in the SPAD estimation model. ② It could be concluded that SPAD-RF regression model, based on the Sentinel – 2 satellite image data, could be used to effectively monitor the SPAD values of leaves of the four vegetation types. Especially for the estimation model of SPAD of reed leaves, R^2 reached 0.926. ③ By analyzing and comparing the model prediction capability under the three schemes, the prediction capability of scheme 3 (including soil parameters) was excellent (2.143 < relative percentage deviation (*RPD*) < 2.692), and the prediction capability was ranked as scheme 3 > scheme 1 > scheme 2. There was a significant nonlinear correlation between the soil properties and the model prediction results. Holistically, Sentinel – 2 data has great potential for predicting chlorophyll content of oasis vegetation. This study provided an efficient, low-cost, potentially high-precision solution to estimate SPAD.

Key words: oasis; Sentinel – 2 data; SPAD; chlorophyll; vegetation index; random forest; Xinjiang